Data Analysis and Statistical Methods
Statistics 651

http://www.stat.tamu.edu/~suhasini/teaching.html

Lecture 21 (MWF) The paired t-test for dependent ‘paired’ samples

Suhasini Subba Rao

Data which does not satisfy the independence result

- Suppose there is a relationship/pairing/matching between X_i and Y_i in the two samples observed.
- When such a relationship exists, then the independence assumption is violated and independent t-test does not give reliable results.
- In such cases, to obtain reliable results we need to do a matched paired t-test, which we motivate and describe below.
- We first consider an example of "matched paired" data and show how applying either the independent t-test (or the Wilcoxon sum rank test) to this data can give rise to unreliable results.

Comparing population, when the independence assumptions are not satisfied: Friday the 13th

- Is Friday the 13th an unusually unlucky day, or is this just another superstition? Does the behaviour of people change on this day?

- Researchers (Scanlon, et al. (1993)) analysed accident patterns on past Friday the 13ths.

- To make the comparision fair they compared the data with what happened the previous week - Friday 6th.

  Why do you think they chose this date to make the comparison?

- Answer to make the fair comparision. They need to compare what happens on the 13th with another date where almost all the factors are the same, except the 13th.

Thursday the 12th is not suitable because, patterns on Thursday could be different to those on Friday. In the UK more accidents happen on Friday and Saturday compared to other days of the week. A completely different time is not fair because more accidents may happen in one season than another. The previous Friday seemed the closest timewise.

The design of an experiment is extremely important!

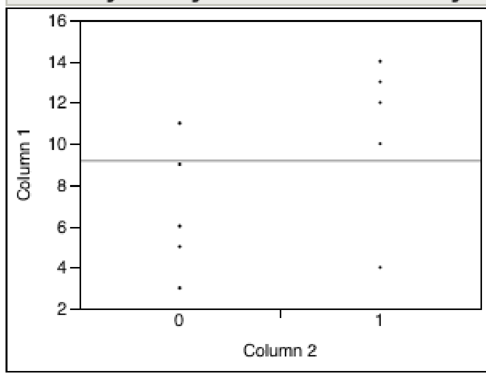
Year	Month	6th	13th	
1989	October	9	13	SWTRHA hospital
1990	July	6	12	SWTRHA hospital
1991	September	11	14	SWTRHA hospital
1991	December	11	10	SWTRHA hospital
1992	March	3	4	SWTRHA hospital
1992	November	5	12	SWTRHA hospital

Friday 13th: Applying the independent t-test

- Let μ_{13} be the mean number of accidents which happen on Friday the 13th and μ_6 the mean number of accidents which happen on Friday the 6th. We want to test (with $\alpha = 0.05$)
  - $H_0 : \mu_{13} - \mu_6 \leq 0$
  - $H_A : \mu_{13} - \mu_6 > 0.$
- The sample mean for the 6th is $\bar{X} = \frac{1}{6}\sum_{i=1}^{6} X_i = 7.5$ and the sample mean for the 13th is $\bar{Y} = \frac{1}{6}\sum_{i=1}^{6} Y_i = 10.83.$

THE JMP output for independent sample t-test

Oneway Analysis of Column 1 By Column 2

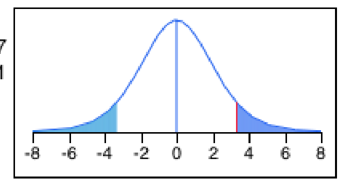


t Test

1-0

Assuming unequal variances

Difference	3.3333	t Ratio	1.664357
Std Err Dif	2.0028	DF	9.940201
Upper CL Dif	7.7994	Prob > |t|	0.1272
Lower CL Dif	-1.1328	Prob > t	0.0636
Confidence	0.95	Prob < t	0.9364



From the JMP output, the p-value for the test is 6.3% $>$ 5%. Thus, at the 5% level, there is not enough evidence to say that the Friday the 13th increases the number of accidents.

- It is believed that peoples behaviour changes on the 13th. But our tests so far do not support this.

- Why not? It could be that we do not have enough data.

  May be, but it could be something more fundamental. Some of the assumptions are being violated and we are not using the correct test.

Did the Friday 13th data satisfy the usual assumptions?

- Let's look at the data again and consider how it was collected.

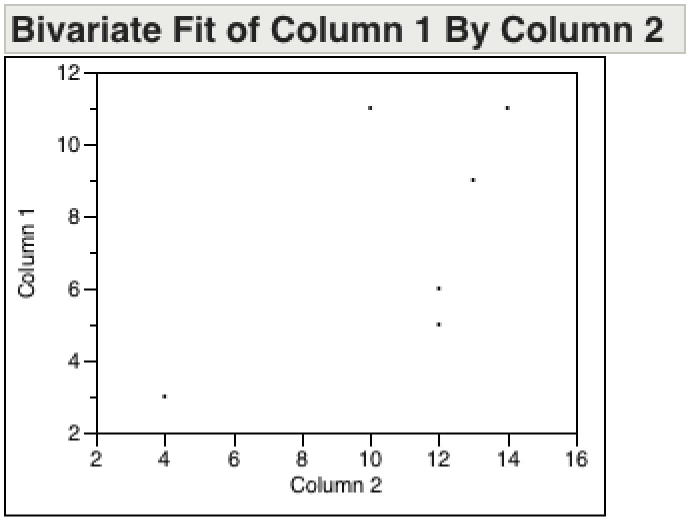
6th	9	6	11	11	3	5
13th	13	12	14	10	4	12
difference	4	6	3	-1	1	7

- We see that all but one of the differences are positive, which suggests by eye that the number of accidents on the 13th are more than those on the 6th.

- Within each sample we can assume data is close to independent, the data was taken at different times of the year. But is each pair independent?

- It is plausible that more accidents happen at certain times of the year. For example, there are more accidents in December than September. This factor drives the dependence between the number of accidents on the successive days.

Plot of Friday 6th against following 13th



6th	9	6	11	11	3	5
13th	13	12	14	10	4	12
diff.	4	6	3	-1	1	7

- When one believes there may be a matching between the data set it is useful to plot it. Often the matching becomes apparent with a "trend".

- Returning to this example, there seems to be a dependence/trend between the number of accidents on the 6th and on the 13th.

This is matched/paired data

- As each 6th of the month can be paired with the following 13th of the month (since they share all factors in common but the date). We called this paired data.

- Looking at the plot, we see that the variation *within* each sample (6th and 13th) is substantial as compared with the differences.

  Any significant difference between the sample means is 'swamped' by the variability in the data.

- It is the issue of the signal and the noise. If the variability is large, rejecting the null is problematic even if the null is true.

- What does this tell us about our assumptions for doing an independent test?

The main thing is that the data on the 6th and the following 13th are not independent.

- Because of dependence between the pairs in the sample, the test as it stands is not appropriate. Instead, we need to do a matched paired t-test.

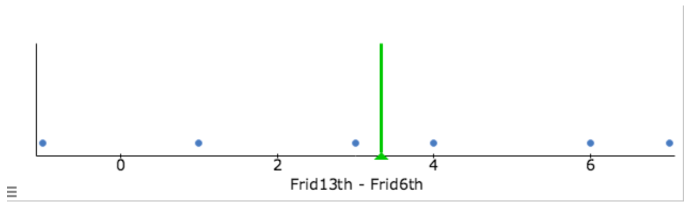
The paired t-test is used when there is dependence within each pair of observations

- When there is a dependence between the pairs take the difference between each pair, and define a new random variable $D_i = X_i - Y_i$.

  The rationale is that by taking the difference, we are reducing the amount of variation which makes it easier to detect a difference on the populations.

- Observe, what happens once the differences are taken in the Friday 13th data. A plot of the differences is also given.

6th	9	6	11	11	3	5
13th	13	12	14	10	4	12
D_i	4	6	3	-1	1	7



- From the plot of the differences, is it unlikely this data could have arisen if there are no differences (or there is a tendency for less accidents to happen on the 13th).

What is precisely happening:

- By considering the difference D_i rather than the individual X_i and Y_i we remove common pairwise factors.

- We treat D_i as a new random variable and simply apply a one sample t-test on the observations $\{D_i\}$.

- If X_i and Y_i come from the same population then the mean of D_i is 0. Let $\mu_d = \mu_X - \mu_Y$ be the population mean of D_i.

The paired t-test

- To test whether the means of μ_X and μ_Y are the same is equivalent to testing $H_0 : \mu_d = \mu_X - \mu_Y \leq 0$ against $H_A : \mu_d = \mu_X - \mu_Y > 0$.

- Now do the 'usual' t-test using the *difference* observations D_i with $(4, 6, 3, -1, 1, 7)$.

- Under the null $\bar{D} \sim N(0, \sqrt{\frac{\sigma^2}{n}})$. Since we have to estimate the standard deviation, we use the regular t-test. Using the data, $\bar{X} = 3.33$ and $s = 3.01$:

$$t_5 = \frac{\bar{D} - 0}{s/\sqrt{n}} = \frac{3.33 - 0}{3.01/\sqrt{6}} = 2.717.$$

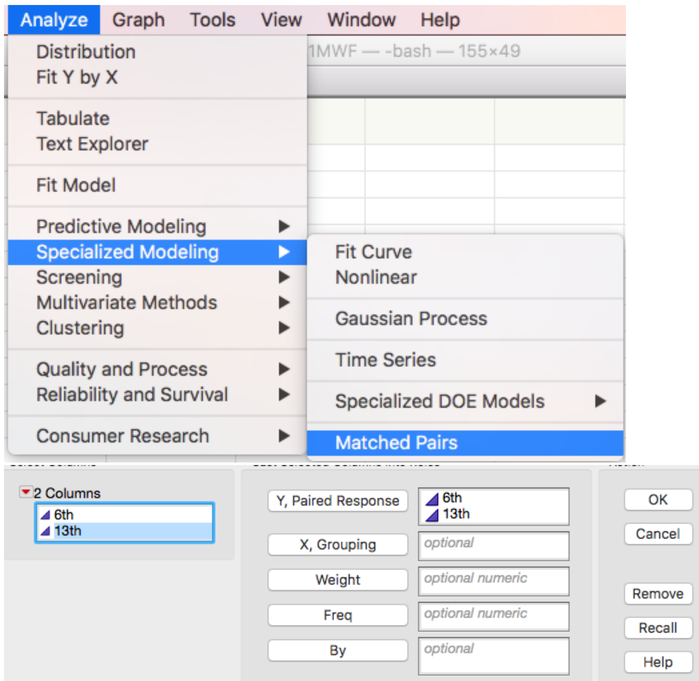
- Since the alternative is pointing write the p-value is $P(t_5 \geq 2.717) = 0.0209$ (we use software to deduce the precise p-value).

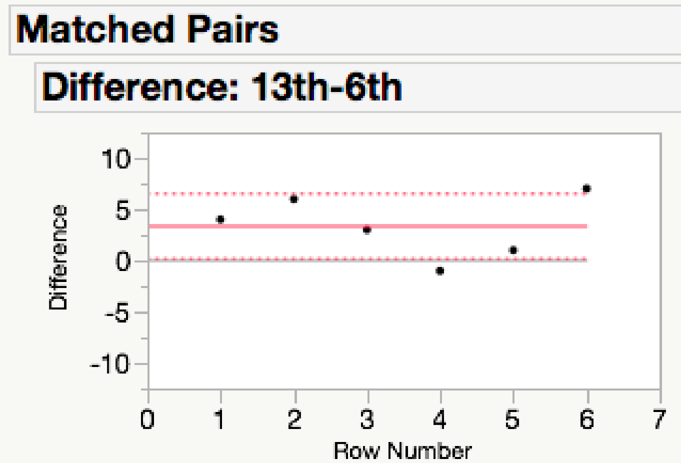
- Since $2.09\% < 5\%$, there is enough evidence to reject the null (at the 5% level). The data suggests that there is an increase in the number of accidents on Friday the 13th as compared with Friday the 6th.

- Of course, this is cumbersome to do by hand. Below we utilize JMP.

- Unlike the independent sample t-test, the matched data must be in two separate columns.

Friday 13th Data: The matched t-test in JMP





13th	10.8333	t-Ratio	2.711631
6th	7.5	DF	5
Mean Difference	3.33333	Prob > |t|	0.0422*
Std Error	1.22927	Prob > t	0.0211*
Upper 95%	6.49328	Prob < t	0.9789
Lower 95%	0.17339		
N	6		
Correlation	0.62515		

Take care which way the difference is entered.

The p-value is 2.1% and we reject the null.

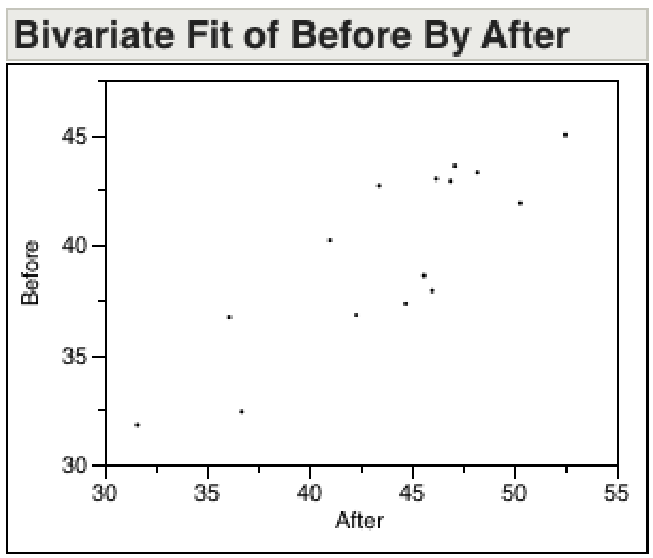
We already came across this test in the Red wine and Polyphenol example!

The paired t-test is very natural. In fact we came across it in the wine and polyphenol example considered in Lecture 16. This was a matched paired t-test in disguise.

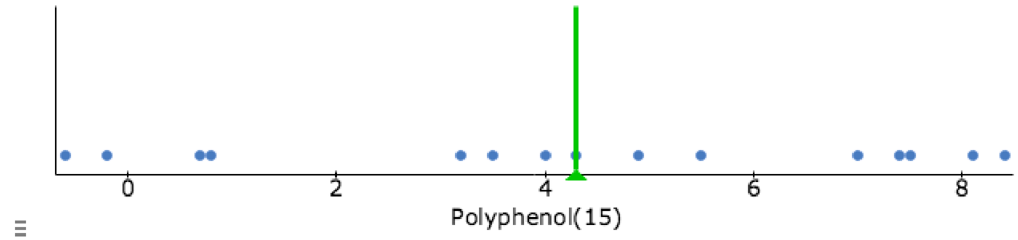
- μ is the population mean *difference* in polyphenol levels before and after taking red wine.
- The data in Lecture 16 had been 'processed', it is the difference in the polyphenol level for each individual, before and after they took the red wine. The unprocessed and pre-processed data looks like this:
- There is a clear 'matching' between the before and after observations (since it is the same individual before and after each examination).

A plot of before against after Wine

To demonstrate that there is matching we plot the before wine against the after wine polyphenol levels. It is clear from there is a linear association between the two data sets and the difference should be taken.



This is a plot of the differences:

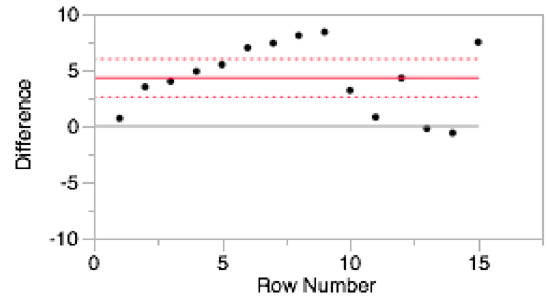


From a plot of the differences, is it likely this data could have arisen if red wine had not positive effect? What do you think the p-value would be for $H_0 : \mu_A - \mu_B \leq 0$ vs. $H_A : \mu_A - \mu_B > 0$?

The paired t-test using the unprocessed polyphenol data

Matched Pairs

Difference: After-Before



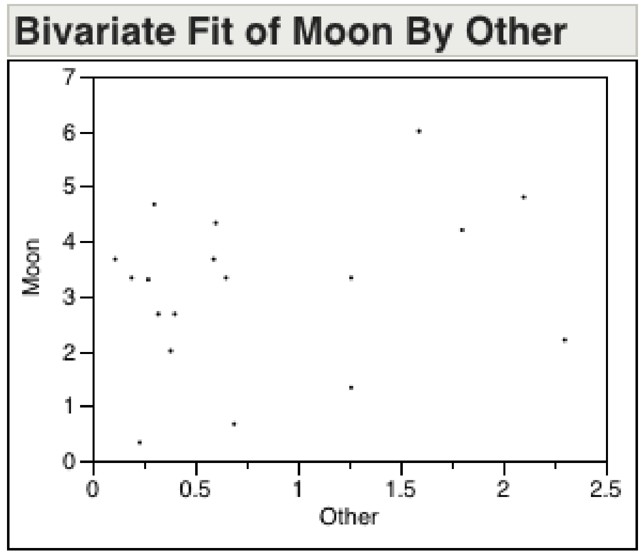
After	43.9067	t-Ratio	5.43188
Before	39.6067	DF	14
Mean Difference	4.3	Prob > |t|	<.0001*
Std Error	0.79162	Prob > t	<.0001*
Upper 95%	5.99786	Prob < t	1.0000
Lower 95%	2.60214		
N	15		
Correlation	0.84566		

The p-value for the test $H_0 : \mu_A - \mu_B \leq 0$ vs. $H_A : \mu_A - \mu_B > 0$ is less than 0.01% and we reject the null. Looking back at the test in Chapter 16, we obtained the same p-value, since the matched paired t-test and one sample t-test on the differences are the same.

The full moon and disruptive patients

A hospital wants to know whether disruptive patients become even more disruptive during the full moon. They monitored disruptive patients over a period of time noting the average number of disruptive instances when there wasn't a full moon and when there was a full moon. The data can be found on my website. There is a clear matching in this data set as the number of disruptive incidences of each patient is monitored.

Plot of moon against non-full moon

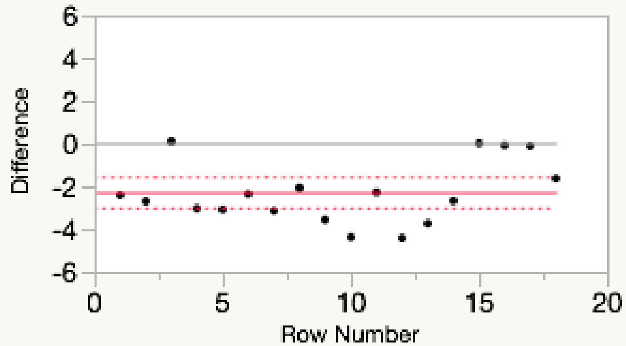


Each point corresponds to an individual patient. There isn't a clear line (there rarely ever is), but we do see some 'evidence' of a linear trend.

Moon JMP output

Matched Pairs

Difference: Other-Moon



Other	0.83556	t-Ratio	-6.71488
Moon	3.13889	DF	17
Mean Difference	-2.3033	Prob > |t|	<.0001*
Std Error	0.34302	Prob > t	1.0000
Upper 95%	-1.5796	Prob < t	<.0001*
Lower 95%	-3.027		
N	18		
Correlation	0.25111		

The Analysis of Full moon data

- We are testing $H_0 : \mu_{NM} - \mu_M \geq 0$ against $H_A : \mu_{NM} - \mu_M < 0$.

- From the output we see that the sample mean difference is -2.3 and this difference is statistically significant with a p-value of less than 0.01%. This means there appears to be more disruptive events during the full moon compared to other times. As this is significant it suggests that more staff should be brought in.

- Next we need to decide on how many extra staff is required, and this depends on the mean number of additional disruptive events. The 95% confidence of $[-3.03, -1.57]$ tells us (with 95% confidence) how many more disruptive events there is likely to be. If you want to be cautious you may take the upper bound and use the upper bound for the mean of 3.03 extra disruptive events per person and calculate the additional staff based on this number. Else you may want to save money and base your calculation on 1.57.

- Of course, this is only a 95% confidence interval, if you want even more confidence you could use a 99% confidence interval $[-2.3 - 2.89 \times 0.34, -2.3 + 2.89 \times 0.34] = [-3.28, -1.31]$.

When to use an independent or a matched paired test

- The independent and matched paired t-test are testing the same hypothesis. The only difference is how the data is collected from the populations.

- If the two samples appear to be completely independent of each other do an independent test

- If there appears a natural 'pairing' in the data do a matched paired test, i.e. the same person before and after a intervention (red wine example, full moon example, running high and low altitude example), the the same month (Friday 13th example), the same bag of M&Ms.

- However, don't be fooled by a spread sheet where there 'appears' to be a natural pairing but isn't.

Example 10 randomly sampled people were put on a diet. 5 were randomly allocated to diet 1 and the other 5 to diet 2. These are two clearly independent samples. However, if the data is collected and displayed in two columns it may give the false impression of pairing, where no real pairing exists.

	Diet 1	Diet 2
1	4.6	3.2
2	2.2	2.6
3	3.1	-1.1
4	6.6	4.4
5	1.2	4.8

- If there logically seems to be a pairing in the data but no clear line is seen in the scatterplot, doing a matched paired t-test still makes sense.

- Your results will still be reliable if you do a matched paired t-test and the samples are completely independent.

- On the other hand, if you do an independent sample t-test when there is matching, the results will not be reliable.